CLAIMS

What is claimed is:

1. A method comprising:

inputting speech representing an utterance and having an intonation; and

identifying an endpoint of the utterance based on the intonation.

2. A method as recited in claim 1, wherein said identifying an endpoint of the utterance based on the intonation comprises comparing the intonation with an intonation model.

3. A method as recited in claim 4, further comprising determining the intonation by computing the fundamental frequency of the utterance.

4. A method as recited in claim 3, wherein said determining the intonation comprises using an intonation model to determine the intonation.

5. A method as recited in claim 1, wherein said identifying the endpoint of the utterance comprises identifying the endpoint of the utterance based on a plurality of knowledge sources, wherein one of the knowledge sources is intonation, including referencing the input speech against a histogram based on training data for each of the knowledge sources.

6. A method as recited in claim 1, further comprising:

determining a period of time that has elapsed since the speech dropped below a threshold value; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the period of time.

7. A method as recited in claim 1, wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on a

length of time for which an energy value of the speech has remained below a predetermined energy value.

8. A method as recited in claim 7, wherein said identifying an endpoint of the utterance further comprises identifying the endpoint of the utterance based on the duration of the final syllable of the utterance.

9. A method of operating an endpoint detector, the method comprising:

inputting speech representing an utterance, the utterance having an intonation; and

comparing the intonation of the utterance with an intonation model;

determining a probability based on a result of said comparing; and

identifying an endpoint of the utterance based on the probability.

10. A method as recited in claim 9, further comprising determining the intonation of the utterance as a function of the fundamental frequency of the utterance.

11. A method as recited in claim 9, further comprising:

determining a period of time that has elapsed since a value of the speech dropped below a threshold value; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the period of time.

12. A method as recited in claim 9, wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the duration of the final syllable of the utterance.

13. A method as recited in claim 12, wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on a

period of time for which an energy value of the speech has remained below a threshold value.

14. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance;

determining that a value of the speech has dropped below a threshold value;

computing an intonation of the utterance;

referencing the intonation of the utterance against an intonation model to determine a first end-of-utterance probability;

determining a period of time that has elapsed since the value of the speech dropped below the threshold value;

referencing the period of time against an elapsed time model to determine a second end-of-utterance probability;

computing an overall end-of-utterance probability as a function of the first and second end-of-utterance probabilities; and

determining whether an end-of-utterance has occurred based on the overall end-of-utterance probability.

15. A method as recited in claim 14, wherein said computing an intonation of the utterance comprises computing an intonation of the utterance by determining the fundamental frequency of the utterance as a function of time.

16. A method as recited in claim 15, further comprising:

determining a duration of a final syllable of the utterance; and

referencing the duration of the final syllable against a syllable duration model to determine a third end-of-utterance probability;

wherein said computing an overall end-of-utterance probability comprises computing the overall end-of-utterance probability as a function of the first, second, and third end-of-utterance probabilities.

17. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance;

computing an intonation of the utterance;

referencing the intonation of the utterance against an intonation model to determine a first end-of-utterance probability;

determining a duration of a final syllable of the utterance;

referencing the duration of the final syllable against a syllable duration model to determine a second end-of-utterance probability;

computing an overall end-of-utterance probability as a function of the first and second end-of-utterance probabilities; and

determining whether an end-of-utterance has occurred based on the overall end-of-utterance probability.

18. A method as recited in claim 17, wherein said computing an intonation of the utterance comprises computing an intonation of the utterance by determining the fundamental frequency of the utterance as a function of time.

19. A method as recited in claim 17, further comprising:

determining that a value of the speech has dropped below a threshold value;

determining a period of time that has elapsed since the value of the speech dropped below the threshold value; and

referencing the period of time against an elapsed time model to determine a second end-of-utterance probability;

wherein said computing an overall end-of-utterance probability comprises computing the overall end-of-utterance probability as a function of the first, second, and third end-of-utterance probabilities.

20. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance, the utterance having a time-varying fundamental frequency;

determining that a value of the speech has dropped below a threshold value;

computing an intonation of the utterance by determining the fundamental frequency of the utterance as a function of time;

referencing the intonation of the utterance against an intonation model to determine a first end-of-utterance probability;

determining a period of time that has elapsed since a value of the speech dropped below the threshold value;

referencing the period of time against an elapsed time model to determine a second end-of-utterance probability;

determining a duration of a final syllable of the utterance;

referencing the duration of the final syllable against a syllable duration model to determine a third end-of-utterance probability;

computing an overall end-of-utterance probability as a function of the first, second, and third end-of-utterance probabilities; and

determining whether an end-of-utterance has occurred by comparing the overall end-of-utterance probability to a threshold probability.

21. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance;

determining an intonation of the utterance;

if the intonation of the utterance is determined to be generally decreasing, then setting a threshold time period equal to a first time value;

if the intonation of the utterance is determined not to be generally decreasing, then setting the threshold time period equal to a second time value larger than the first time value; and

identifying an endpoint of the utterance based on the threshold time period.

22. A method as recited in claim 21, wherein said using the threshold time period to identify an endpoint of the utterance comprises using the threshold time period to identify an endpoint of the utterance by determining that an endpoint of the utterance has occurred if an energy value of the speech remains below a predetermined value for the threshold time period.

23. A method as recited in claim 21, wherein said determining an intonation of the utterance comprises using an intonation model.

24. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance, the utterance having a time-varying fundamental frequency;

determining an intonation of the utterance by

computing the intonation as the fundamental frequency of the utterance as a function of time, and

referencing the intonation against an intonation model to determine the intonation of the utterance;

if the intonation of the utterance is determined to be generally decreasing, then setting a threshold time period equal to a first time value;

if the intonation of the utterance is determined not to be generally decreasing, then setting the threshold time period equal to a second time value larger than the first time value; and

using the threshold time period to identify an endpoint of the utterance, by determining that an endpoint of the utterance has occurred if the speech remains below a predetermined value for a length of time equal to the threshold time period.

25. A machine-readable program storage medium tangibly embodying a sequence of instructions executable by a machine to perform a method for endpoint detection, the method comprising:

inputting speech representing an utterance, the utterance having an intonation; and

identifying an endpoint of the utterance based on the intonation of the utterance.

26. A machine-readable program storage medium as recited in claim 25, wherein said using the intonation of the utterance in identifying an endpoint of the utterance comprises comparing the intonation of the utterance with an intonation model.

27. A machine-readable program storage medium as recited in claim 25, wherein the method further comprises determining the intonation of the utterance.

28. A machine-readable program storage medium as recited in claim 27, wherein said determining the intonation of the utterance comprises computing the fundamental frequency of the utterance.

29. A machine-readable program storage medium as recited in claim 27, wherein said determining the intonation of the utterance comprises using an intonation model to determine the intonation of the utterance.

30. A machine-readable program storage medium as recited in claim 25, wherein the method further comprises:

determining a period of time for which an energy value of the speech has been below a threshold value; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the period of time.

31. A machine-readable program storage medium as recited in claim 25, wherein the method further comprises:

determining a duration of a final syllable of the utterance; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the duration of a final syllable of the utterance.

32. A machine-readable program storage medium as recited in claim 31, wherein the method further comprises:

determining a period of time that has elapsed since a value of the speech dropped below a threshold value; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the period of time.

33. An endpoint detector comprising:

means for inputting speech representing an utterance, the utterance having an intonation; and

means for identifying an endpoint of the utterance based on the intonation of the utterance.

34. An endpoint detector as recited in claim 33, wherein said means for using the intonation of the utterance in identifying an endpoint of the utterance comprises means for comparing the intonation of the utterance with an intonation model.

35. An endpoint detector as recited in claim 33, further comprising means for determining the intonation of the utterance.

36. An endpoint detector as recited in claim 35, wherein said means for determining the intonation of the utterance comprises means for computing the fundamental frequency of the utterance.

37. An endpoint detector as recited in claim 35, wherein said means for determining the intonation of the utterance comprises means for using an intonation model to determine the intonation of the utterance.

38. An endpoint detector as recited in claim 33, further comprising:

means for determining a period of time that has elapsed since a value of the speech dropped below a threshold value; and

wherein said means for identifying an endpoint of the utterance comprises means for identifying the endpoint of the utterance further based on the period of time.

39. An endpoint detector as recited in claim 33, further comprising:

means for determining a duration of a final syllable of the utterance; and

wherein said means for identifying an endpoint of the utterance comprises means for identifying the endpoint of the utterance further based on the duration of a final syllable of the utterance.

40. An endpoint detector as recited in claim 39, further comprising:

means for determining a period of time that has elapsed since a value of the speech dropped below a threshold value; and

wherein said means for identifying an endpoint of the utterance comprises means for identifying the endpoint of the utterance further based on the period of time.

41. An apparatus for performing endpoint detection comprising:

means for inputting speech representing an utterance, the utterance having a time-varying fundamental frequency;

means for determining that a value of the speech has dropped below a threshold value;

means for computing an intonation of the utterance by determining the fundamental frequency of the utterance as a function of time;

means for referencing the intonation of the utterance against an intonation model to determine a first end-of-utterance probability;

means for determining a period of time that has elapsed since the speech dropped below the threshold value;

means for referencing the period of time against an elapsed time model to determine a second end-of-utterance probability;

means for referencing the duration of the final syllable of the utterance against a syllable duration model to determine a third end-of-utterance probability;

means for computing an overall end-of-utterance probability as a function of the first, second, and third end-of-utterance probabilities; and

means for determining whether an end-of-utterance has occurred by comparing the overall end-of-utterance probability to a threshold probability.